

Controlling the number of focal elements

Some combinatorial considerations

Christophe OSSWALD

Abstract A basic belief assignment can have up to 2^n focal elements, and combining them with a simple conjunctive operator will need $\mathcal{O}(2^{2n})$ operations. This article proposes some techniques to limit the size of the focal sets of the bbas to be combined while preserving a large part of the information they carry.

The first section revisits some well-known definitions with an algorithmic point of view. The second section proposes a matrix way of building the least committed isopignistic, and extends it to some other bodies of evidence. The third section adapts the k -means algorithm for an unsupervised clustering of the focal elements of a given bba.

Key words: Basic belief assignments, Combinatorial complexity, Focal elements, k -means, Pignistic probability, Body of evidence, Least commitment

1 General considerations on basic belief assignments

Let the finite set $X = \{x_1, \dots, x_n\}$ be our frame of discernment. The size of X will be noted $n = |X|$. The set of all the subsets of X will be noted 2^X .

Definition 1. [Shafer(1976)] The application m from 2^X to $[0, 1]$ is a *basic belief assignment* (bba) if :

$$\sum_{A \subseteq X} m(A) = 1 \quad (1)$$

The constraint of *closed world* is modeled by $m(\emptyset) = 0$. If $m(\emptyset)$ is greater than 0, we either have an *open world* or a *conflict* within the information.

Christophe OSSWALD

ENSTA Bretagne, Lab-STICC UMR 3192 e-mail: Christophe.Osswald@ensta-bretagne.fr

Definition 2. Let m be a bba on X . $A \subseteq X$ is a *focal element* of m if $m(A) > 0$. The *focal set* of m is composed of all its focal elements :

$$F(m) = \{A \subseteq X \mid m(A) > 0\} \quad (2)$$

The size of m is noted $|m| = \text{Card}(F(m))$.

Of course, $|m| \leq 2^n$. In most applications, $|m|$ will be very small compared to 2^n when a bba is constructed from a source's information, but after some steps of combination, this limit can be reached.

Definition 3. Let m be a bba on X . The most usual *bodies of evidence* are :

- The *belief*:

$$\text{bel}(A) = \sum_{\substack{B \subseteq A, \\ B \neq \emptyset}} m(B) = \sum_{\substack{B \subseteq A, \\ B \neq \emptyset, \\ B \in F(m)}} m(B) \quad (3)$$

- The *plausibility*:

$$\text{pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) = \sum_{\substack{B \cap A \neq \emptyset, \\ B \in F(m)}} m(B) \quad (4)$$

- The *commonality*:

$$\text{q}(A) = \sum_{B \supseteq A} m(B) = \sum_{\substack{B \supseteq A, \\ B \in F(m)}} m(B) \quad (5)$$

- The *pignistic probability*, which is additive (knowing $\text{betP}(\{x\})$ for all $x \in X$ is sufficient):

$$\text{betP}(A) = \frac{1}{1 - m(\emptyset)} \sum_{B \subseteq X} \frac{|A \cap B|}{|B|} m(B) = \frac{1}{1 - m(\emptyset)} \sum_{B \in F(m)} \frac{|A \cap B|}{|B|} m(B) \quad (6)$$

When the context is not obvious, the bba used to define the body of evidence will be placed as an index : $\text{betP}_m(A)$ instead of $\text{betP}(m)$.

In the definition 3, the first expression concerns all the subsets of X , and the second expression concerns only the focal elements. Therefore, if f is either of the bodies of evidence, and A a subset of X , a natural implementation of the equation brings an algorithm which calculates $f(A)$ in $\mathcal{O}(2^n)$ operations with the first expression. As the second expression only browses the focal set of m , its complexity is $\mathcal{O}(|m|)$, for the same result.

The most popular combination operator is the non-normalized conjunctive rule, also known as Smet's rule. It is a quite simple operator to implement; it is associative, and therefore allows to combine many sources.

Definition 4. Let m_1 and m_2 be two bbas on X . The *conjunctive combination* of m_1 and m_2 is a bba on X , $m_1 \oplus m_2$, defined by :

$$(m_1 \oplus m_2)(A) = \sum_{\substack{B \subseteq X, \\ C \subseteq X, \\ B \cap C = A}} m_1(B)m_2(C) = \sum_{\substack{B \in F(m_1), \\ C \in F(m_2), \\ B \cap C = A}} m_1(B)m_2(C) \quad (7)$$

The cost for calculating $B \cap C$ is $\mathcal{O}(n)$. The first expression brings an algorithm in $\mathcal{O}(n2^{2n})$ operations for calculating $(m_1 \oplus m_2)(A)$, and $\mathcal{O}(n2^{3n})$ for determining $m_1 \oplus m_2$. The second expression brings an algorithm in $\mathcal{O}(n|m_1||m_2|)$ operations for calculating $(m_1 \oplus m_2)(A) = (m_1 \oplus m_2)(A)$, and $\mathcal{O}(n2^n|m_1||m_2|)$ for determining $m_1 \oplus m_2$.

Smets [Smets(2002)] proposed a nice implementation in $\mathcal{O}(n2^n)$ operations for transformations between bba and commonality. The conjunctive combination of the commonality functions is a simple multiplication, which is linear, but on vectors having a size of 2^n .

The expression (7), nor the commonality, can prevent us from making operations on non-focal elements of $m_1 \oplus m_2$. Let the bba be implemented by an adaptive structure that contains information only for its focal elements. A hashtable is a convenient way for it. The algorithm 1 uses only $\mathcal{O}(n|m_1||m_2|)$ to build $m_1 \oplus m_2$.

The size of m_\cap is at most $|m_1||m_2|$. The algorithm coming from (7) needs to be executed for all the subsets of X , but the algorithm 1 only works on the focal elements of m_\cap , and does not compute useless intersections [Smets(1994)]. Using a hashtable for the focal elements, with a hashcode calculation in $\mathcal{O}(n)$ operations, the conjunctive combination takes $\mathcal{O}(n|m_1||m_2|)$ operations.

```

Data: bbas  $m_1, m_2$ 
Result: bba  $m_\cap$ 
forall  $B \in m_1$  do
  forall  $C \in m_2$  do
    if  $B \cap C \in m_\cap$  then
       $m_\cap(B \cap C) \leftarrow m_\cap(B \cap C) + m_1(B)m_2(C)$ 
    else
      Add  $B \cap C$  to  $m_\cap$ 
       $m_\cap(B \cap C) \leftarrow m_1(B)m_2(C)$ 

```

Algorithm 1: Conjunctive combination

However, the very nature of the combination operator brings a combinatorial explosion of the focal set. Let m_i be the bba defined by $m_i(X) = \frac{1}{2}$ and $m_i(X \setminus \{x_i\}) = \frac{1}{2}$: $|m_i| = 2$. Let m_\cap be the conjunctive combination of all those bbas : $m_\cap = m_1 \oplus \dots \oplus m_n$. For any $A \subseteq X$, $m_\cap(A) = \frac{1}{2^n}$. Therefore, $F(m_\cap) = 2^X$ and $|m_\cap| = 2^n$.

The objective of the following sections will be to guarantee that the size of a bba cannot be too large, and to respect its nature as much as possible.

2 Linear algebra for bbas

The definition 3 builds the bodies of evidence bel , pl , betP and q as linear transformations of m . Considering a bba m on X and an integer K , our objective will be to build an bba m' on X such that $|m'| \leq K$ and $f_{m'}(A) = f_m(A)$ for some bodies of evidence f and some subsets A of X .

Within this section, we forbid \emptyset to be a focal element of m , and we do not allow it to become a focal element of m' . As convenient consequences, we have $\text{bel}(A) \leq \text{betP}(A) \leq \text{pl}(A)$, $\text{bel}(X) = 1$, and $\text{pl}(X) = 1$.

A popular and efficient way to build a bba from a probability or another source of uncertain information is to build a least committed bba having the same pignistic probability than the source [Smets(1990)].

Definition 5. Let m be a bba on X . A bba m' is an *isopignistic* of m if

$$\forall x \in X, \text{betP}_m(x) = \text{betP}_{m'}(x) \quad (8)$$

The bba m' is the *least committed isopignistic* of m if for any isopignistic m'' of m and for any $A \subseteq X$, $\text{pl}_{m'}(A) \geq \text{pl}_{m''}(A)$.

The algorithm 2 builds the least committed isopignistic in $\mathcal{O}(n^2 + n|m|)$ operations. It contains at most n focal elements.

Data: bba m on X
Result: bba m' on X
forall $x \in X$ **do**
 Calculate $p[i] = \text{betP}(x)$
 $A \leftarrow X$; $k \leftarrow |X|$
while $\max(p) \neq 0$ **do**
 $i \leftarrow \text{argmin}(p)$
 $m'(A) \leftarrow kp[i]$
 forall $j \in p$ **do**
 $p[j] \leftarrow p[j] - p[i]$
 Delete element i from p
 $A \leftarrow A \setminus \{x_i\}$; $k \leftarrow k - 1$

Algorithm 2: Building the least committed isopignistic

If we calculate $\text{betP}(x)$ for all $x \in X$, and order the elements of X such that $p_i = \text{betP}(x_i) \geq \text{betP}(x_{i+1}) = p_{i+1}$, the focal elements of the least committed isopignistic are a subset of the $A_i = \{x_1, \dots, x_i\}$.

We have

$$p_i = \text{betP}(x_i) = \sum_{k=i}^n \frac{1}{k} m'(A_k) \quad (9)$$

Let p be the vector of the p_i and y be the vector of the $m'(A_i)$. We have $p = \text{Bet}y$ with Bet a $n \times n$ matrix, triangular and inversible. Therefore

$y = \text{Bet}^{-1}p$, with

$$\text{Bet} = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ 0 & \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ \vdots & 0 & \frac{1}{3} & \cdots & \frac{1}{n-1} & \frac{1}{n} \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \frac{1}{n-1} & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \frac{1}{n} \end{pmatrix}, \text{Bet}^{-1} = \begin{pmatrix} 1 & -1 & 0 & \cdots & \cdots & 0 \\ 0 & 2 & -2 & 0 & & \vdots \\ 0 & 0 & 3 & -3 & \ddots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & 0 \\ \vdots & & & \ddots & (n-1) & -(n-1) \\ 0 & \cdots & \cdots & \cdots & 0 & n \end{pmatrix} \quad (10)$$

As Bet^{-1} is a triangular band matrix, we can compute all the $m'(A_i)$ from p_i in $\mathcal{O}(n)$ operations.

With $\mathcal{O}(n|m|)$ operations for computing betP , $\mathcal{O}(n \ln n)$ operations for sorting X , $\mathcal{O}(n)$ operations for building the sets A_i (with an adapted data structure) and $\mathcal{O}(n)$ operations for solving the linear system, building the least committed isopignistic costs $\mathcal{O}(n(\ln n + |m|))$ operations. Usually, $|m| \gg \ln n$, and the cost of the least committed isopignistic is not greater than the cost of computing $\text{betP}(x)$ for the elements of X .

The interval $[\text{bel}(A), \text{pl}(A)]$, containing $\text{betP}(A)$, can be interpreted as an uncertainty on A [Janez and Appriou(1996)]. For singletons, bel is trivial: $\text{bel}(x) = m(x)$. For sets of size $n - 1$, pl is trivial: $\text{pl}(X \setminus \{x\}) = 1 - m(\{x\})$. Considering the non-trivial bodies of evidence on the sets of interest $\{x_1\}, \dots, \{x_n\}$, $B_1 = X \setminus \{x_1\}, \dots, B_n = X \setminus \{x_n\}$, we search a bba m' with those focal elements, forming a vector

$$\mathbf{y} = (m'(\{x_1\}), \dots, m'(\{x_n\}), m'(B_1), \dots, m'(B_n))^T \quad (11)$$

which verifies:

$$\forall i \in [1, n], \text{pl}_{m'}(\{x_i\}) = \text{pl}_m(\{x_i\}) \quad (12)$$

$$\forall i \in [1, n], \text{bel}_{m'}(B_i) = \text{bel}_m(B_i) \quad (13)$$

We have:

$$\text{pl}_{m'}(\{x_i\}) = m'(\{x_i\}) + \sum_{j \neq i} m'(B_j) \quad (14)$$

$$\text{bel}_{m'}(B_i) = \sum_{j \neq i} m'(\{x_j\}) + m'(B_i) \quad (15)$$

As $\forall i, \text{pl}_{m'}(\{x_i\}) + \text{bel}_{m'}(B_i) = \sum_i m'(\{x_i\}) + \sum_i m'(B_j)$, there are only $n + 1$ independent equations among the $2n$ listed above: we cannot guarantee to keep at the same time $\text{pl}_m(\{x_i\})$ and $\text{bel}_m(B_i)$ on those $2n$ focal elements.

As $q(B_i) = m(B_i) + m(X)$ and $q(\{x_i\}) = \text{pl}(\{x_i\})$, introducing commonality does not bring any new independent equation.

2.1 Mixing Bet with other bodies of evidence

Here we search a bba with $2n$ focal elements which is an isopignistic of m and respects an other body of evidence on some focal elements. In the following examples, we allow the A_i obtained in section 2 to be focal elements, and we complete them with $(\{x_i\})_{i \in [1, n]}$ or the $(B_i)_{i \in [1, n]}$.

With plausibility, we should use the focal elements $(\{x_i\})_{i \in [1, n]}$. We build a vector

$$\mathbf{y} = (m'(\{x_1\}), \dots, m'(\{x_n\}), m'(A_1), \dots, m'(A_n))^T \quad (16)$$

The constraints are:

$$\text{betP}(x_i) = m'(\{x_i\}) + \sum_{k=i}^n \frac{1}{k} m'(A_k) \quad (17)$$

$$\text{pl}(\{x_i\}) = m'(\{x_i\}) + \sum_{k=i}^n m'(A_k) \quad (18)$$

As $A_1 = \{x_1\}$, we cannot have $m'(A_1) \neq m'(\{x_1\})$; we have only $2n - 1$ focal elements. We drop the term $m'(\{x_1\})$ in \mathbf{y} , and the constraint on $\text{pl}(\{x_i\})$ to obtain a matrix P such that

$$P\mathbf{y} = (\text{pl}_m(\{x_2\}), \dots, \text{pl}_m(\{x_n\}), \text{betP}(x_1), \dots, \text{betP}(x_n))^T \quad (19)$$

The matrix P_4 and more generally P_n are:

$$P_4 = \begin{pmatrix} 0 & 0 & 0 & 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ 1 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ 0 & 1 & 0 & 0 & 0 & \frac{1}{3} & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 & 0 & \frac{1}{4} \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}, \quad P_n = \left(\begin{array}{c|c} 0 \dots 0 & \\ \hline I_{n-1} & \text{Bet}_n \\ \hline I_{n-1} & \begin{array}{c} 0 \\ \vdots \\ 0 \end{array} \end{array} \middle| \begin{array}{c} \\ \\ U_{n-1} \end{array} \right) \quad (20)$$

where Bet_n is matrix obtained in the section 2 and U_{n-1} the upper triangular $(n-1) \times (n-1)$ matrix full of 1.

The matrix P_n is invertible, and we can solve this system in $\mathcal{O}(n^3)$ operations. Overall, we can reduce the focal set of m to $2n - 1$ focal elements in $\mathcal{O}(n(n^2 + |m|))$ operations, respecting betP and pl on the singletons.

With commonality, we obtain the same results : $q(\{x_i\}) = \text{pl}(\{x_i\})$.

With belief, we should use $(B_i)_{i \in [1, n]}$ as focal elements instead of $(\{x_i\})$. As $\text{bel}(B_i) + \text{pl}(\{x_i\}) = 1$, we obtain another – but similar – $(2n-1) \times (2n-1)$ invertible matrix.

3 Optimatization by k -means

[Denoeux and Yaghlane(2002)] proposed to reduce a bba by adapting the single linkage hierarchical clustering algorithm to coarsen its focal set. Another interesting family of unsupervised clustering algorithm are the k -means techniques, born from the ISODATA method of [Ball and Hall(1965)]. One can adapt this method to find a subset \mathcal{K} of 2^X limited in size: $|\mathcal{K}| \leq k$.

Usual k -means does not guarantee an optimal choice of centers: finding them is equivalent to the MINIMUM- k CENTER, which is a NP-Complete problem [Garey and Johnson(1979)]. The convergence of the k -means algorithm is guaranteed, but only to a local minimum of the intra-cluster variance.

Data: bba m , integer k with $k \leq |m|$
Result: bba m_k
 Let $C[1], \dots, C[k]$ be k focal elements of m [1]
repeat
 forall $j \leq k$ **do** $\mathcal{C}[j] \leftarrow \emptyset$
 forall $A \in m$ **do** [2]
 $\mathcal{C}[\text{argmin}(\text{dist}(A, C_j)))] \leftarrow \mathcal{C}[\text{argmin}(\text{dist}(A, C_j))] \cup \{A\}$
 forall $j \leq k$ **do** $C[j] \leftarrow \text{center of } \mathcal{C}[j];$ [3]
until *ending condition reached* [4]
forall $j \leq k$ **do**
 $m_k(C[j]) \leftarrow \sum_{A \in \mathcal{C}[j]} m(A)$

Algorithm 3: k -means, in a general way that applies to focal elements.

[1] It is natural to initialize the algorithm with the k focal elements with the greatest masses. But, as the algorithm converges – if it converges – to a local minimum, it should be a good idea to execute various instances, with random starting sets.

[2] The focal element A is affected to the center $C[j]$ such that

$$\text{dist}(A, C[j]) = \left| \left(A \cap \overline{\mathcal{C}[j]} \right) \cup \left(\overline{A} \cap C[j] \right) \right| \quad (21)$$

is minimal. It corresponds to a natural L_1 distance based on an exclusive OR. In case of equal distances to different centers, it is possible to:

- choose a random one *(the algorithm is no longer deterministic)*
- use a lexicographical order *(elements are no longer equivalent)*
- try to build balanced clusters *(the underlying problem is NP-complete)*

[3] The usual k -means technique uses the geometrical barycenter of the focal sets of $\mathcal{C}[j]$ seen as points of $[0, 1]^n$: $C[j] \leftarrow \sum_{A \in \mathcal{C}[j]} m(A)A$.

It would build fuzzy focal elements, which is not the way the definition 2 accepts them. Therefore, we put x in the new $C[j]$ if and only if :

$$\sum_{A \in \mathcal{C}[j], x \in A} m(A) > \sum_{A \in \mathcal{C}[j], x \notin A} m(A) \quad (22)$$

- [4] As we “move” the centers of the classes to the nearest sharp subset of X , the total intra-cluster variance is not necessarily decreasing. Therefore, the ending condition must include a maximum steps number, and/or test the cycles it should encounter.

A step of the algorithm 3 costs $\mathcal{O}(kn|m|)$ operations. A reasonable number of steps before ending the loop is k , and we obtain an algorithm in $\mathcal{O}(k^2n|m|)$ operations. If we want to compare this approach with the ones of the section 2.1, we should use $k = 2n - 1$, and get an algorithm in $\mathcal{O}(n^3|m|)$ operations.

4 Conclusion

In a general way, dealing with basic belief assignments on large frames of discernment need a proper encoding of the focal sets. We propose to use hashtables for this purpose, but this not the only way. We propose two categories of methods for restricting any bba to a bba modest in focal set size.

We extend the principle of isopignistic to other bodies of evidence to build a bba with only $2n-1$ focal elements, respecting both the pignistic probability and another body of evidence of the original bba. We first determine the value of the bodies of evidence on some simple elements, and then determine the restricted focal set. A linear equation gives the restricted bba.

Trying to restrict the focal set to a number of representative elements leads to a NP-Complete problem. We adapt the k -mean algorithm to build a heuristical solution. It is more expensive, but it does not need to define *a priori* a focal set, and can adapt to more situations.

References

- Ball and Hall(1965). Ball GH, Hall DJ (1965) Isodata, a novel method of data analysis and pattern classification. Tech. rep., Stanford Research Institute
- Denoeux and Yaghlane(2002). Denoeux T, Yaghlane AB (2002) Approximating the combination of belief functions using the fast moebius transform in a coarsened frame. International Journal of Approximate Reasoning 31(1-2):77-101
- Garey and Johnson(1979). Garey MR, Johnson DS (1979) Computers and intractability – a guide to the theory of NP-Completeness. Freeman
- Janez and Appriou(1996). Janez F, Appriou A (1996) Théorie de l’Evidence et cadres de discernement non exhaustifs. Traitement du Signal 13(3):237-250
- Shafer(1976). Shafer G (1976) A mathematical theory of evidence. Princeton University Press
- Smets(1990). Smets P (1990) Constructing the pignistic probability function in a context of uncertainty. Uncertainty in Artificial Intelligence 5:29-39
- Smets(1994). Smets P (1994) The transferable belief model. Artificial Intelligence 66:191-234
- Smets(2002). Smets P (2002) Matrix Calculs for Belief Functions. International Journal of Approximate Reasoning 31:1-30